TRANSFORMTIKA, Vol.18, No.1, July 2020, pp. 93 - 101
ISSN: 1693-3656,
journals.usm.ac.id/index.php/transformatika

Pengelompokkan Data Akademik Menggunakan Algoritma K-Means Pada Data Akademik Unissula

Dedy Kurniadi[1], Andre Sugiyono[2]
[1]Universitas Islam Sultan Agung/Jurusan Teknik Informatika
Jl. Raya Kaligawe KM.4, telp: 024-6583584, e-mail: ddy.kurniadi@unissula.ac.id
[2]Universitas Islam Sultan Agung/Jurusan Teknik Industri
Jl. Raya Kaligawe KM.4, telp: 024-6583584, e-mail: andre@unissula.ac.id

ARTICLE INFO

Article history:

Received 15 May 2020
Received in revised form 15 June 2020
Accepted 16 June 2020
Available online 31 July 2020

ABSTRACT

A Higher education in the digital era as it is now common to use IT technology (Information Technology) in supporting their daily activities, but the use of IT raises a problem that is serious enough if there is no support and no further management this is only produce a data noise , Sultan Agung Islamic University (Unissula) has implemented an IT-based academic information system, in use of this information systems by time this systems produce a lot of data in the unissula academic database and this data is monotonous data and not clustered or also called data noise data that overlap without any benefit and information further in it, the purpose of this study is to solve the problem of these data into student performance data based on the GPA from semester 1 to semester 4 and make it to be a best data to support an alternative decision by the leader, this study uses the method of datamining and k-means algorithms, k-means algorithm is very good to be used as a solution for problems related to *clustering*, k-means algorithm is an algorithm that is unsupervised and the data can be adjusted by its self according to its class, the results of this study are a decision support system for grouping academic data for *clustering* student that have good potential and *drop out* potential in the form of dashboard information systems.

Keywords: k-means Algorithm; Academic Database; Information Systems

1. Introduction

Pada era digital seperti sekarang sebuah institusi perguruan tinggi dapat dipastikan memiliki sebuah sistem informasi akademik sebagai media untuk pelaporan kegiatan yang ada pada perguruan tinggi tersebut, efek dari munculnya sistem tersebut adalah terciptanya banyak

Received May 15, 2020; Revised June 15, 2020; Accepted June 16, 2020

data, artinya semakin bertambahnya data pada sebuah institusi lambat laun data tersebut akan menjadi besar.

Universitas Islam Sultan Agung (Unissula) sudah mengimplementasikan Sistem Informasi dengan banyak modul namun sistem informasi di Unissula masih menggunakan metode transaksional hal tersebut menjadikan data yang ada hanya tersimpan didalam database dan pengelola kesulitan dalam memperoleh laporan tentang data akdemik yang valid karena data hanya dimasukkan dan dibaca saja tanpa ada pengolahan menggunakan algoritma didalamnya, untuk mengolah data tersebut agar lebih terstruktur dibutuhkan sebuah metode datamining.

Datamining merupakan metode untuk menemukan berbagai informasi yang baru dan penting didalam tumpukan data, diperlukan *datawarehouse* untuk mengolah data-data yang ada pada database, *datawarehouse* adalah teknik yang mengoleksi data. yang mempunyai sifat time-variant, non-volatile dan subject-oriented yang mampu memproses *Bigdata* [1], di Unissula terdapat data yang sangat besar dan perlu diklasifikasi menjadi status akademik, algoritma untuk klasifikasi yang digunakan adalah K-Means *Clustering*, algoritma k-means mampu memproses data dan membaginya menjadi beberapa kelompok dan dimasukkan ke dalam cluster, algoritma k-means menerima inputan berbentuk vektor dan datanya mengelompokkan sendiri-sendiri sesuai dengan kelasnya tanpa target yang ditentukan algoritma ini sangat cocok untuk proses klasifikasi status akademik mahasiswa [2].

Sistem akademik sendiri merupakan sistem untuk memonitoring pelaporan akademik mahasiswa tiap periode, bentuk pelaporan di bagian informasi akademik adalah status mahasiswa, aktifitas kuliah mahasiswa, data kelas kuliah, nilai akademik mahasiswa, masa studi dan overstudi, sistem akademik yang sekarang belum mencakup pelaporan-pelaporan terkait overstudi, pengelompokkan mahasiswa non aktif terancam *drop out* dan pengelompokan mahasiswa aktif non aktif dengan IPK dibawah peraturan akademik yang terancam *drop out*, lulus tepat waktu atau lebih cepat dengan banyaknya data maka hal tersebut tidak bisa diselesaikan dengan cara manual oleh karena itu diperlukan sebuah sistem informasi untuk menyelesaikan permasalahan tersebut, sistem informasi akademik dengan banyak data akan diolah dengan memanfaatkan metode datamining menggunakan algoritma K-Means untuk menemukan informasi – informasi baru yang nantinya berguna bagi pengambilan keputusan oleh pemangku kebijakan.

Data yang dihasilkan dari penelitian ini degenerate secara otomatis menggunakan sistem untuk mendapatkan hasil yang maksimal data akademik di klasterisasi menjadi tiga klaster yaitu terancam lulus diluar waktu normal, lulus diwaktu normal dan lulus lebih cepat dari waktu normal dan juga mendeteksi mahasiswa dengan status non aktif disetiap semester dengan filter menandai data tersebut apabila lebih dari 2 kali non aktif selama menempuh perkuliah di Universitas Islam Sultan Agung Semarang.

2. Research Method

2.1 Data *Clustering*

Data *Clustering* adalah jenis metode dalam Data Mining dengan sifat tak terarah (*unsupervised*). Terdapat dua jenis data *clustering* yang kerap digunakan dalam proses pengelompokan data pertama adalah jenis hirarki *clustering* dan yang kedua non hirarki data *clustering*. K-Means adalah salah satu teknik yang menggunakan data *clustering* non hirarki bekerja dengan cara mempartisi data yang ada ke dalam satu bentuk atau lebih cluster/kelompok. Metode k-means mempartisi data ke dalam cluster/kelompok dengan data yang memiliki karakteristik yang sama kemudian dikelompokkan ke dalam satu cluster data, kemudian data yang mempunyai karakteristik berbeda akan dikelompokkan dalam kelompok cluster data yang lain. Tujuan dari data *clustering* untuk meminimalisasikan *objective function* data yang diset dalam proses *clustering* dan memaksimalisasikan variasi antar cluster [3], [4].

2.2 Perkembangan K-Means

Beberapa alternatif penerapan K-Means dengan beberapa pengembangan teori-teori penghitungan terkait telah diusulkan. Hal ini termasuk pemilihan :

1. Distance space untuk menghitung jarak di antara suatu data dan centroid
2. Metode pengalokasian data kembali ke dalam setiap cluster
3. *Objective function* yang digunakan

Distance space adalah jarak antara data dan centroid, jarak antara dua titik X1 dan X2 pada block distance space dihitung dengan menggunakan rumus sebagai berikut [5], [6]:

$$\mathrm{DL1\ (X1, X2)} = || \mathrm{X2 - X1} || 1 = \sum_{j=1}^{p} | \mathrm{X2j - X1j} | \quad (1)$$

Dimana :

P = Dimensi Data

| . | = Nilai Absolut

2.3 Metode Fuzzy K-Means

Metode Fuzzy K-Means mengalokasikan data ke dalam cluster data masing-masing dengan memanfaatkan teori Fuzzy. metode Fuzzy K-Means menggunakan variabel membership function, Uik , yang menentukan rujukan seberapa besar kemungkinan data bisa dimasukkan menjadi anggota ke dalam suatu cluster data. fuzzy k-menas juga menggunakan variabel m yang merupakan weighting exponent dari membership function. Variabel m bisa mengubah nilai besaran pengaruh membership function, Uik , pada proses *clustering* dengan menggunakan metode Fuzzy K-Means. m memiliki area nilai m>1. Nilai m yang sering dan umum digunakan adalah 2. Membership function untuk suatu data ke suatu cluster tertentu dihitung menggunakan rumus sebagai berikut [7] [8]:

$$\mathrm{Uik} = \sum_{j=1}^{e} \left(\frac{D(Xk,Vi)}{D(Xk,Vj)} \right)^{\frac{2}{m-1}} \quad (2)$$

Dimana :

Uik = Membership function data ke-k ke cluster ke-i

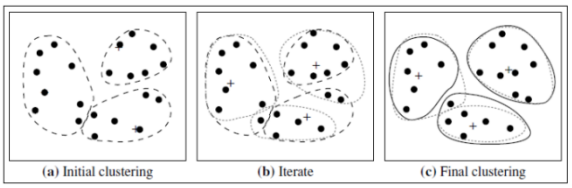
Vi = Nilai centroid cluster ke-i

m = Weighting Exponent

2.4 Teknik *Clustering*

Teknik *Clustering* merupakan bentuk analisa yang digunakan untuk mengidentifukasi cluster-cluster yang terdapat pada data yang digunakan, cluster sendiri merupakan sebuah object data yang sifatnya mirip atau sama antara cluster satu dan cluster lainnya. Teknik *clustering* sendiri sering digunakan untuk membuat data menjadi terstruktur shingga data tersebut bermanfaat, proses *clustering* melakukan proses ekstrasi dari data yang tidak diketahui kemudian mencari pattern yang tersembunyi dari data yag besar [9]. Tujuan dilakukannya *clustering* adalah untuk menyortir objek-objek atau data-data yang berbeda kedalam satu kelompok atau grup (cluster) dimana derajat asosiasinya memiliki nilai yang maksimal apabila dikelompokkan menjadi satu grup.

Data yang sudah disusun dan direpresentasikan kedalam kelompok-kelompok dengan nilai yang maksimal kemudian akan direpresentasikan sebagai suatu kelompok dan membentuk sebuah subset sebagai contoh C=C1,…..,Ck dari S, sehingga S= ∪i=1k Ci dan $Ci \cap Cj$= $\emptyset$ untuk i ≠ j, dari subset tersebut data pada S yang masuk kedalam satu subset *Clustering* merupakan data-data dengan subset yang memiliki karakteristik yang sama [10].



Gambar 2. 1 K-Means *Clustering* [10]

2.5 Algoritma K-Means

Teknik K-Means merupakan teknik yang sederhana namun betul-betul maksimal dalam proses pengolahan datanya sehingga bisa didapat data yang valid, tahapan K-Means adalah menentukan nilai K terlebih dahulu sebagai pusat centroid, nilai awal K digunakan sebagai parameter yang ditentukan secara manual. Kemudian setiap data dijumlahkan dengan semua centroid dan dihitung jarak antar cluster (distance) yang akan membentuk nilai cluster dari setiap data dan akan dikelompokkan secara random, kemudian tahapan-tahapan tadi dilakukan kembali dengan centroid baru disetiap perulangannya sampai dengan centroid tersebut tidak ada perubahan maka data cluster tersebut dinyatakan valid (Singh, et al, 2013).

2.6 Eucledian Distance

Euclidian distance adalah jarak antar data satu dengan data yang lainnya dalam satu baris data. *Euclidean distance* melakukan perhitungan akar antar data untuk mendapatkan perbedaan antara objek yang berpasangan (Fahim, et al, 2006). Tujuan dari dilakukannya perhitungan jarak antar data ini adalah sebuah proses k-means *clustering* untuk penentuan data yang dihitung akan masuk ke cluster yang mana.

$$\mathrm{dis}((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2} \quad (3)$$

Dimana :

x,y = titik satu data

a,b = titik data lainnya

3. Results and Analysis

Data yang sudah didapat diimplementasi dengan menggunakan perhitungan algoritma k-means, data tersebut adalah seperti berikut.

Tabel 3.1 Data Mahasiswa

No	Nama	IPK 1	IPK 2	IPK 3	IPK 4
1	maria ulfa	3.76	3.68	3.67	3.66
2	agung dwi prasetyo	0.9	0.9	0.9	0.9
3	ahmad al hasan	3.68	3.45	3.51	3.56
4	ahmad jalaluddin	3.18	3.1	3.11	3.23
5	almas adlil wafi	0	0	0	0
6	cita setiyo putri	2.78	2.91	3.04	3.15
7	firmansyah	2.15	2.38	2.41	2.41
8	ilham mahmud	2.08	2.32	2.32	2.32
9	irfan nur chaerul	2.93	2.96	2.78	2.78
10	johan hamuda	2.75	2.66	2.68	2.82

No	Nama	IPK 1	IPK 2	IPK 3	IPK 4
11	johan yudha aditya	2.83	2.74	2.79	2.89
12	m. nur ihsan	0.35	0.35	0.35	0.35
13	muchamad danial	2.83	2.71	2.68	2.82
14	m. dhonny mahendra	2.63	2.63	2.74	2.8
15	muhammad thariq alvin	3.13	3.14	3.11	3.32
16	muhammad vicky budiyono	3.33	3.15	3.15	3.3
17	muh ridwan syarif alghifari	1.25	1.51	1.51	1.95
18	muzaroah	3.28	3.28	3.29	3.45
19	rafika arini rahmawati	3	2.84	2.83	3.03
20	rizky miftachul huda	2.9	2.93	2.91	3.03
21	rizqi firmansyah	3.13	2.81	2.8	3
22	sakinah	3.38	3.26	3.37	3.52
23	sigit ardianto	3.55	3.16	3.24	3.42
24	soufi maulani muttafaq	2.83	2.75	2.73	2.83
25	vina amanatul maula	0.9	0.9	0.9	0.9
26	yessy ambar dewi	0	0	0	0
27	yusuf arief wicaksono	3.3	3.09	3.11	3.28
28	achmad maryadi	3.13	2.83	2.93	3.06
29	agus putra	2.18	1.7	2	2.14
30	mohamad ekky putra perdana	3.4	3.15	3.19	3.16
31	noor hidayanto	3.19	3.27	3.27	3.1
32	kukuh ramadhan sukma putra	3.33	2.72	2.72	2.72

4.1 Pra Proses

Selanjutnya data mrntah tersebut dihitung centroid awalnya dan ditentukan cluster dengan jarak terdekatnya untuk pembentukan cluster awal dan sebagai acuan pembentukan cluster selanjutnya sampai dengan tidak adanya perubahan pada cluster dan centroid yang telah diproses.

Rumus yang digunakan adalah

$$\mathrm{dis}((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2}$$

Tabel 3.2 Data Iterasi I

No	Nama	dc1	dc2	dc3	C
1	maria ulfa	2.88	1.06	0.66	3
2	agung dwi prasetyo	2.73	4.55	4.97	1
3	ahmad al hasan	2.61	0.78	0.38	3
4	ahmad jalaluddin	1.81	0.11	0.47	2
5	almas adlil wafi	4.52	6.35	6.77	1
6	cita setiyo putri	1.43	0.46	0.85	2
7	firmansyah	0.16	1.69	2.11	1
8	ilham mahmud	0.00	1.84	2.26	1

No	Nama	dc1	dc2	dc3	C
9	irfan nur chaerul	1.25	0.69	1.09	2
10	johan hamuda	0.97	0.90	1.31	2
11	johan yudha aditya	1.13	0.73	1.14	2
12	m. nur ihsan	3.83	5.65	6.07	1
13	muchamad danial	1.05	0.84	1.25	2
14	muchammad dhonny mahendra	0.90	0.96	1.37	1
15	muhammad thariq alvin	1.84	0.00	0.43	2
16	muhammad vicky budiyono	1.98	0.21	0.33	2
17	muh ridwan syarif alghifari	1.46	3.26	3.68	1
18	muzaroah	2.14	0.30	0.15	3
19	rafika arini rahmawati	1.37	0.52	0.92	2
20	rizky miftachul huda	1.38	0.47	0.89	2
21	rizqi firmansyah	1.43	0.55	0.93	2
22	sakinah	2.26	0.43	0.00	3
23	sigit ardianto	2.22	0.45	0.26	3
24	soufi maulani muttafaq	1.08	0.79	1.20	2
25	vina amanatul maula	2.73	4.55	4.97	1
26	yessy ambar dewi	4.52	6.35	6.77	1
27	yusuf arief wicaksono	1.90	0.18	0.40	2
28	achmad maryadi	1.51	0.44	0.81	2
29	agus putra	0.73	2.37	2.77	1
30	mohamad ekky putra perdana	1.97	0.32	0.42	2
31	noor hidayanto	1.91	0.31	0.47	2
32	kukuh ramadhan sukma putra	1.43	0.85	1.16	2

Dari iterasi pertama sampai dengan iterasi terakhir harus dihitung distance masing-masing data sehingga menghasilkan data valid.

Tabel 3.3 Data Iterasi ke n

No	Nama	dc1	dc2	dc3	C
1	maria ulfa	2.88	1.06	1.06	3
2	agung dwi prasetyo	2.73	4.55	4.55	1
3	ahmad al hasan	2.61	0.78	0.78	3
4	ahmad jalaluddin	1.81	0.11	0.11	2
5	almas adlil wafi	4.52	6.35	6.35	1
6	cita setiyo putri	1.43	0.46	0.46	2
7	firmansyah	0.16	1.69	1.69	1
8	ilham mahmud	0.00	1.84	1.84	1
9	irfan nur chaerul	1.25	0.69	0.69	2
10	johan hamuda	0.97	0.90	0.90	2

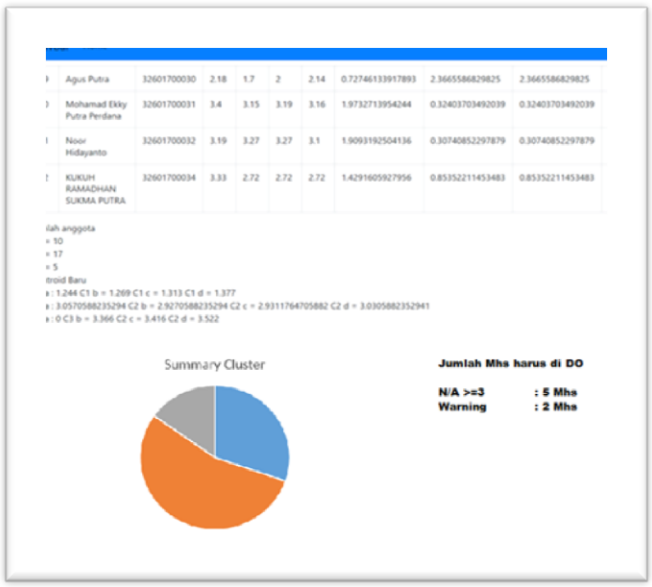
No	Nama	dc1	dc2	dc3	C
11	johan yudha aditya	1.13	0.73	0.73	2
12	m. nur ihsan	3.83	5.65	5.65	1
13	muchamad danial	1.05	0.84	0.84	2
14	muchammad dhonny mahendra	0.90	0.96	0.96	1
15	muhammad thariq alvin	1.84	0.00	0.00	2
16	muhammad vicky budiyono	1.98	0.21	0.21	2
17	muh ridwan syarif alghifari	1.46	3.26	3.26	1
18	muzaroah	2.14	0.30	0.30	3
19	rafika arini rahmawati	1.37	0.52	0.52	2
20	rizky miftachul huda	1.38	0.47	0.47	2
21	rizqi firmansyah	1.43	0.55	0.55	2
22	sakinah	2.26	0.43	0.43	3
23	sigit ardianto	2.22	0.45	0.45	3
24	soufi maulani muttafaq	1.08	0.79	0.79	2
25	vina amanatul maula	2.73	4.55	4.55	1
26	yessy ambar dewi	4.52	6.35	6.35	1
27	yusuf arief wicaksono	1.90	0.18	0.18	2
28	achmad maryadi	1.51	0.44	0.44	2
29	agus putra	0.73	2.37	2.37	1
30	mohamad ekky putra perdana	1.97	0.32	0.32	2
31	noor hidayanto	1.91	0.31	0.31	2
32	kukuh ramadhan sukma putra	1.43	0.85	0.85	2

Dilakukan sampai dengan distance tidak berubah, kemudian didapatkan hasil yang paling valid yaitu iterasi terakhir dimana distance sebelumnya dan distance terakhir sama, data pada centroid dan data cluster dengan centroid

C1 a : 1.244, C1 b = 1.269, C1 c = 1.313, C1 d = 1.377,
C2 a : 3.0570588235294 C2 b = 2.9270588235294
C2 c = 2.9311764705882 C2 d = 3.0305882352941
C3 a : 0 C3 b = 3.366 C2 c = 3.416 C2 d = 3.522

Jumlah anggota C1 = 10; C2 = 17; C3 = 5

Hasil keseluruhan ditunjukkan dalam dashboard sistem pada gambar 4.1.



Gambar 3.1 Dashboard Sistem Informasi

4. Conclusion

Berdasarkan penelitian yang sudah dilakukan terdapat kesimpulan yang bisa diambil, sistem pendukung keputusan pengelompokkan data akademik ini berhasil diterapkan ke dalam sistem informasi akademik Universitas Islam Sultan Agung Semarang dengan model sistem informasi prototype, metode data mining menggunakan algoritma K-Means berhasil diterapkan dan menghasilkan cluster dengan masing anggota pada cluster 1 terdapat jumlah anggota sebanyak 31% pada cluster 2 terdapat jumlah anggota sebanyak 53% dan cluster 3 terdapat jumlah anggota sebanyak 16%, sistem pendukug keputusan ini juga mampu menampilkan jumlah mahasiswa yang terancam *Drop out* (DO) dan jumlah mahasiswa yang peru diperingatkan (*warning*) dengan melihat data akademik yang ada.

References

[1] Sutrisno, Afriyudi, and Widiyanto, "Penerapan Data Mining Pada Penjualan Menggunakan Metode *Clustering* Study Kasus Pt . Indomarco," *Penerapan Data Min. Pada Penjualan Menggunakan Metod. Clust.*, vol. Vol.x No.x, no. Data Mining, pp. 1–11, 2013.

[2] X. Wu and R. Srihari, "New ν-support vector machines and their sequential minimization\r\nalgorithm," *Twent. Int. Conf. Mach. Learn.*, 2003.

[3] Yudi Agusta, "K-Means – Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3, no. 11, pp. 47–60, 2007.

[4] J. O. Ong, "Implementasi Algoritma K-Means *Clustering* Untuk Menentukan Strategi Marketing," *J. Ilm. Tek. Ind.*, vol. 12, no. 1, pp. 10–20, 2013.

[5] M. Kaur and S. Kang, "Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 78–85, 2016.

[6] Y. C. Chen and C. T. Su, "Distance-based margin support vector machine for classification," *Appl. Math. Comput.*, vol. 283, pp. 141–152, 2016.
[7] H. Jia, "Large-Scale Data Classification Method Based on Machine Learning Model," *Int. J. Database Theory Appl.*, vol. 8, no. 2, pp. 71–80, 2015.
[8] T. Li, Y. Chen, X. Mu, and M. Yang, "An improved fuzzy k-means *clustering* with k-center initialization," *3rd Int. Work. Adv. Comput. Intell. IWACI 2010*, vol. 1009, pp. 157–161, 2010.
[9] H. Islam and M. Haque, "An Approach of Improving Student's Academic Performance by using K-means *clustering* algorithm and Decision tree," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 8, pp. 146–149, 2012.
[10] Y. Prastyo, "Pembagian Tingkat Kecanduan Game Online Menggunakan K-Means *Clustering* Serta Korelasinya Terhadap Prestasi Akademik," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 2, no. 2, p. 138, 2017.
[11] A. Singh, A. Yadav, and A. Rana, "K-means with Three different Distance Metrics," *Int. J. Comput. Appl.*, vol. 67, no. 10, pp. 13–17, 2013.
[12] A. M. Fahim, A. M. Salem, F. A. Torkey, and M. A. Ramadan, "Efficient enhanced k-means *clustering* algorithm," *J. Zhejiang Univ. Sci.*, vol. 7, no. 10, pp. 1626–1633, 2006.